THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 52, No. 8, October, 1973
Printed in U.S.A.

Picture Coding: The Use of a Viewer Model in Source Encoding*

By J. O. LIMB

(Manuscript received March 22, 1973)

A method is suggested for inserting viewer criteria directly into coding algorithms; any complex visual model may be used. The technique is applied to a DPCM-type coder, and a number of variations are compared on the basis of entropy, quality, and complexity. It is found that, using a simple one-dimensional filter model, the first-order entropy of the DPCM signal can be reduced by 30 percent for a high-detail picture with only a small reduction in picture quality. Furthermore, by means of a single threshold control, one can efficiently trade off bit-rate and picture quality over a large range for use in adaptive strategies.

I. INTRODUCTION

In early work in picture coding, Graham stressed the role of the viewer and Powers and Staras concluded that if large reductions in bit-rate are to be achieved they must come from "nonstatistical" (perceptual) redundancies.[1,2] However, there have been few attempts to explicitly incorporate the viewer in the encoder design. Unfortunately, there is no general method for handling complex viewer fidelity criteria, especially when one is concerned with how pleasing a picture appears.† Nevertheless *ad hoc* techniques have been proposed and evaluated and have achieved a certain measure of success.[3–8]

Source encoding, in its most general form, can be diagrammed as shown in Fig. 1. The first stage is an irreversible operation which generates a discrete signal as a result of a quite general multidimensional quantization process. The resulting discrete signal may still be redundant due to the presence of statistical dependencies; these are removed in the second stage of reversible processing in which a digital

* Presented, in part, at the 1972 IEEE International Symposium on Information Theory.

† See Ref. 9 for a discussion on viewer fidelity criteria.

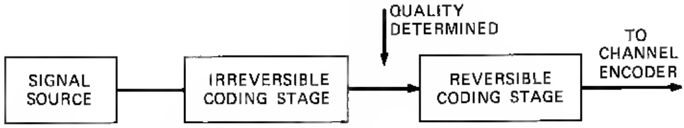

Fig. 1—General source-encoding model.

sequence is assigned to the output of the first stage. Thus, in the first stage the properties of the receiver together with the signal statistics are incorporated into the quantizing process so that the resulting signal just meets the required quality.

At the output of the first stage, picture quality is established and a discrete entropy can be measured. The actual transmission rate will then approach the entropy depending on how well the second encoding stage is designed to fit the statistics of the source.

1.1 *Receiver-Model Coding*

Algorithm: Components of a picture signal are estimated by some method. A test is made to see whether the estimate is adequate by testing the estimate on a model of the receiver. If so, the receiver is told (implicitly or explicitly) that the estimate is adequate. If not, a component is transmitted so as to meet the required criterion.

This type of algorithm will be referred to as "receiver-model" coding. Obviously, it is a rather general approach which can be appended to a larger number of existing algorithms; for example, the interpolators and predictors summarized by Kortman.[10] In this study we are interested in applying it to the differential quantizer (DPCM coder) although even here it can be applied in many ways.

In designing a coder to incorporate properties of the human observer the most important subjective effect is probably the large decrease in visual sensitivity that occurs adjacent to a change in luminance.[11–13] An attempt to design a coder based on this effect leads to some form of the familiar differential quantizer (DPCM coder).[3,7,8]

Probably the second most important subjective effect is the change in visual sensitivity with average luminance (Weber effect).[14] However, in the television situation nonlinearity between applied voltage and output luminance in most displays partially offsets this change in sensitivity, so that, roughly speaking, noise on an electrical television signal is nearly equally visible throughout the luminance range.[15,16]

Probably the third most important subjective effect is the spatial filtering of small-amplitude, luminance perturbations. It is this third

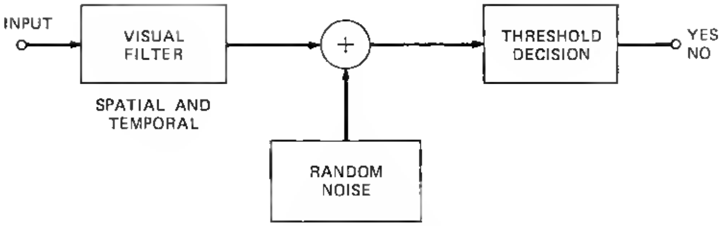

Fig. 2—Simple model of visual threshold filtering.

receiver property which we will attempt to capitalize on through the use of receiver-model coding in this paper.

A simple model that is reasonably successful in explaining the visibility of liminal stimuli is shown in Fig. 2.[13,17,18] Because we are dealing with very small perturbations (at least at the neural level) we will ignore nonlinearities. The input stimulus on which the model is developed is here a small luminance perturbation on a uniform background. The stimulus undergoes temporal and spatial filtering and in the process is corrupted by noise, represented as an additive random component. The filtered signal with the perturbation is compared with the filtered background signal. If the difference exceeds a certain threshold then the perturbation will be visible.* The model is quite accurate for variously shaped stimuli presented on a uniform background with the exception that if the stimulus is long (subtended angle >1 degree) and thin (subtended angle <5 minutes), it will be significantly more visible than the model predicts.[19,13]

The situation is more complex in the case of normal picture evaluation. First the perturbation is not presented against a uniform background and second the perturbation is not directly presented to the viewer; instead it is the difference between the coded picture and the viewer's memory of the original. Thus, although we will use this particular filter model it should be upgraded as we understand more about the visibility of perturbations in a complex scene.

In this study we will only be concerned with the spatial effect of the visual filter; different shapes have been postulated for the spatial impulse response and in one study the Gaussian function was found to fit as well as any.[13]† However, as we shall see, the performance of the algorithm is not sensitive to the exact shape that is used. The degree of spread, compared with the size of a picture element, is shown

* Because of the linearity assumption it does not matter if we filter the difference (error) signal or filter the two signals separately and then subtract.

† See also recent work of Ref. 20.

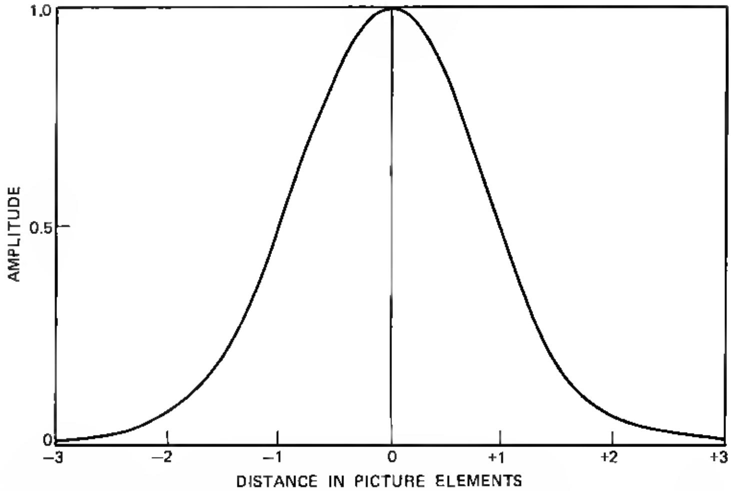

Fig. 3—Spatial impulse response of vision: visual point-spread function for a *Picturephone*®-type display reviewed at 36 inches.

in Fig. 3 for a *Picturephone*®-type display at a standard viewing distance of 36 inches. One should note that this filter is only appropriate to threshold vision; once a perturbation is much above threshold it may no longer be applicable.

Note that the efficacy of the filtering operation depends very much on viewing distance. Thus, one would expect that at smaller viewing distances the eliminated components would no longer be subliminal while at larger viewing distances the threshold filtering process could be taken further.

1.2 *Coding Algorithms*

Receiver-model coding will be applied to the differential quantizer by means of an interpolative algorithm.[10] Consider that sample i (Fig. 4) is the last nonzero sample that has been quantized and that sample $i + j$ is now being processed. The difference $X_{i+j} - \hat{X}_i$ is formed (where $\hat{X}_i$ is the differentially quantized value of X_i), it is quantized, and the discrete value of X_{i+j}, $\hat{X}_{i+j}$ is evaluated (i.e., normal differential quantizer operation). Interpolated values of the intermediate samples $\bar{X}_{i+1}, \cdots, \bar{X}_{i+j-1}$ are then formed from $\hat{X}_i$ and $\hat{X}_{i+j}$ and the error sequence associated with the interpolated values is calculated. This error sequence is than processed by the filter-threshold circuit to determine whether the errors are visually acceptable or not. If the error sequence associated with sample $i + j$

passes the test, the algorithm steps to sample $i + j + 1$ and no new value is transmitted. If tbe test fails, the run is terminated, tbat is, the quantized difference associated with sample $i + j - 1$ is transmitted.

There are two distinct forms which the coding algorithm may take; free-running or grid. In the free-running algorithm a maximum length-of-run is specified in advance for practical reasons. If the interpolation attempts to continue beyond the maximum length, a new sample is taken and a new run commenced. In most studies the maximum length-of-run is 10 pels. In the grid algorithm a fixed set of pels (grid elements) is always transmitted and interpolation or extrapolation is only applied to the intervening elements. Fixed patterns corresponding to every second or every fourth element along a line have been studied and the pattern is offset (staggered) from line to line. Grid algorithms are studied because in some forms they are very much simpler to implement.

Section II gives the experimental details and describes the basis for comparing different algorithms while Section III describes the operation and performance of a free-running interpolative algorithm and explores the effects of error filtering. In Section IV we describe and compare the operation of a number of different grid algorithms including one which involves but a minor modification to the normal differential quantizer.

II. EXPERIMENTAL ARRANGEMENT

The different algorithms were evaluated using a computer facility. The 8-bit digitized pictures are read from a digital disk, line at a time,

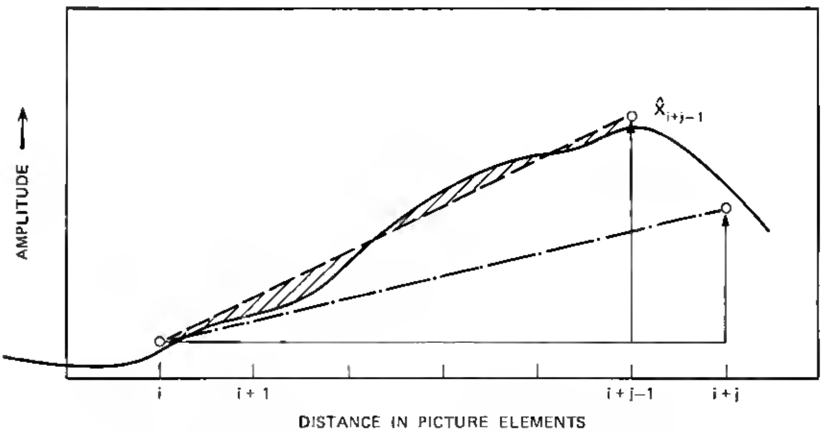

Fig. 4—In description of an extrapolative threshold coder.

processed, and then stored in a digital frame store for direct viewing on a television monitor. The picture consists of 250 lines with 210 elements in each line. The picture is generated and displayed as a 2:1 interlaced picture at 30 frames (60 fields) per second; hence adjacent lines in the picture originate in different fields. This format is similar to the *Picturephone* format.

In evaluating the picture we look at a single frame, repeated at 30 frames per second; thus temporal effects are not considered. The picture quality is slightly hetter when viewing a "frozen" frame of a differentially quantized picture since "edge husyness" and certain random noise components are noticeahly less ohjectionahle in the frozen situation contrary to the findings for white noise.[21]

2.1 *Differential Quantizer*

The normal differential quantizer is the vehicle with which the various algorithms will he tested. The 13-level companded quantizing characteristic is given in Tahle I. The differential quantizer has no integrator "leak" hut the integrator is reset at the heginning of each line.

The results will he given mainly in terms of two different pictures. The first picture is the familiar "Karen" which hy most measures would he regarded as active and is fairly difficult to code if both the soft hair and the sharp stripes are to he preserved. The second picture is much simpler having a large flat hackground and is referred to as

TABLE I—QUANTIZER CHARACTERISTIC OF 13-LEVEL DIFFERENTIAL QUANTIZER
(expressed in 1/128ths of the $p - p$ amplitude)

Level Number	Decision Level	Representative Level
0		0
	1	
±1		2
	3	
±2		4
	6	
±3		8
	11	
±4		14
	18	
±5		22
	27	
±6		32

"Lamp." A third picture ("Birdcage") is occasionally used; it is intermediate in complexity hetween the previous two.

The picture quality of the differentially quantized signal is only distinguishahle from the 8-hit digital signal hy careful comparison; there is a slight increase in hackground noise and very small amounts of slope overload and edge-husyness. The discrete, first-order entropies of the three pictures after coding hy the differential quantizer are 3.10, 2.79, and 2.37 hits/pel for Karen, Birdcage, and Lamp, respectively. The second-order entropies are 2.92, 2.61, and 2.20 hits/pel, respectively. Thus, little would he gained in the second stage of coding, the reversihle stage, hy any attempt to remove higher-order redundancy.

2.2 *Quality*

One difficulty in documenting the performance of coders lies in specifying the quality of the processed pictures.

One can divide picture quality into different ranges hy using a set of criteria. Consider the following three:

1. Difference just detectahle hy a skilled ohserver hetween the processed and unprocessed pictures in an A–B comparison with no restriction on viewing distance.
2. Defects just noticeable to a skilled ohserver at standard viewing distance (36 inches approximately $7H$) for a picture with which the ohserver is *familiar*.
3. [*] Defects just noticeahle to a skilled ohserver, at standard viewing distance when the ohserver has *no knowledge* of the original picture.

The picture quality of criterion 1 is prohahly the most frequently used *ad hoc* criterion hut it is unnecessarily severe for visual communication purposes and, if employed, would result in a significant increase in hit-rate over that required hy criteria 2 and 3. In this study the author has attempted to specify the qualities of coded pictures using criteria 2 and 3. This is inevitahly an approximate process and as a consequence a range is given rather then a specific value. Approximate as this process is, if it enahles a rank ordering of coding strategies it will have served its purpose.

---

[*] Where the viewer was familiar with the test picture a conscious effort was made to disregard defects that depend on knowledge of the original picture. For example, noticing a loss of fine detail in the hair region of "Karen" depends on memory of the original; noticeable slope overload on the other hand generally appears as an unnatural distortion.

2.3 *Bit-Rate Calculation*

Picture quality and hit-rate are the two vital measures of coder performance. In this study we are concerned primarily with the first coding stage of Fig. 1, the multidimensional quantizing stage. But the final hit-rate will also depend on how thoroughly the second stage is implemented. However, what we will do is to calculate entropies of the signal after the first stage of coding, the rationale heing that the figure represents a hound on what is ohtainahle in practice. In some instances variahle wordlength coding, with huffering, will yield a data rate that is within a few percent of the entropy figure.[22,23] In other instances more complex coding will he required to approach the entropy figures, particularly for source alphahets which contain a highly prohahle event where something akin to runlength coding would be required.

The performance of the algorithms has heen assessed hy calculating the entropy under the assumption of two different types of reversible encoding. They are:*

Code I. All pels in the run are processed in the same way (with the same code). This is the simplest hut most inefficient method. The hit-rate hound is ohtained hy calculating the first-order entropy of the signal;

$$H_1 = -\sum_{i=1}^{N} p_i \log p_i,$$

where p_i is the prohahility of occurrence of each event (a quantizer level or an interpolate command) and N is the total numher of different types of event.

Code II. A separate code is used for each run position. That is, the first element in a run uses code 1, the second element in the run uses code 2, etc. The run is terminated hy the sampled pel. The entropy is then given hy

$$H_2 = \sum_{j=1}^{M} h_j q_j,$$

where h_j is the entropy of events in the jth position of the run and q_j is the prohahility that an event will he in the jth position of a run.

* In some of the algorithms to be discussed the information indicating that an element has been successfully estimated at the transmitter is ohtained indirectly hy the receiver from the coded bit stream. In other algorithms an additional code word is appended for this purpose.

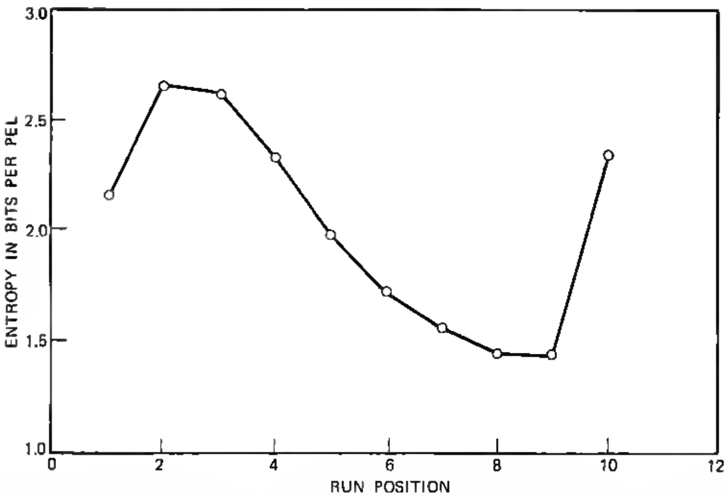

Fig. 5—Variation of entropy with the position of the element in the run; free-running interpolative algorithm with a maximum runlength of 10. Subject—Karen.

The entropy of the signal changes significantly depending on the position in the run. (This is shown in Fig. 5 where the first-order entropy of the differentially quantized signal is plotted as a function of the position in the run for a free-running interpolative algorithm having a maximum runlength of 10.) This change in entropy is exploited in code II (but not code I). Where the average length of a run is large, a practical realization of a code II coder could well result in a type of runlength encoding.

In summary, H_1 can be regarded as the lower bound on data rate when each element is coded in the same way while H_2 is a lower bound when run contiguity is exploited.

There are a great number of different techniques for reversibly coding the discrete output of the first coding stage (Fig. 1); by specifying the abovementioned two entropies we can concentrate more on the irreversible stage without getting overly involved in exactly how the second-stage coding will be achieved. The entropies are always given as bits/active (or unblanked) picture element.

III. RESULTS: FREE-RUNNING ALGORITHM

The details of the interpolative algorithm are summarized in the flow diagram of Fig. 6. Bookkeeping operations like entering a new line, testing for the end of a line, and gathering statistics are not

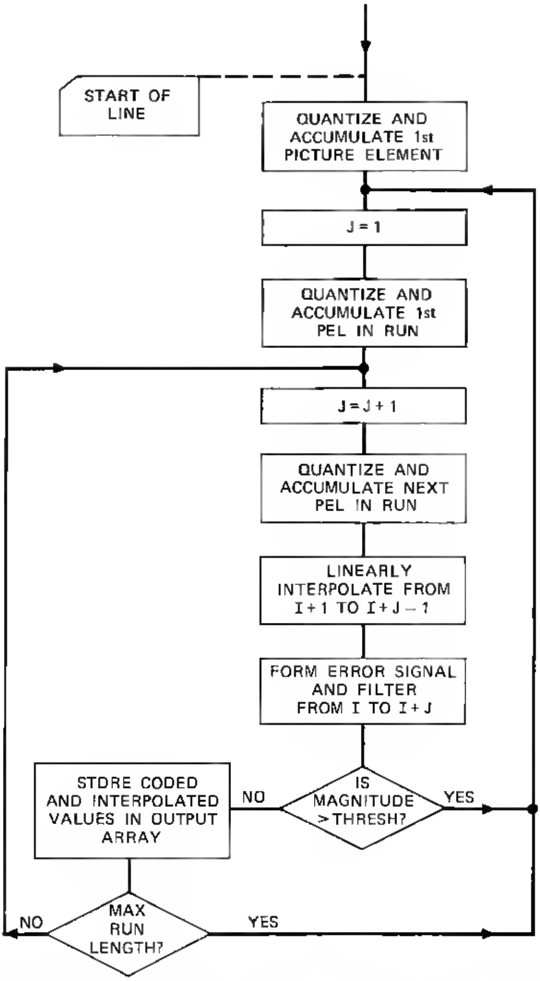

Fig. 6—Flow diagram for the element processing of the free-running interpolative algorithm. I denotes the last element in the previous run, J denotes the current length of the run being processed, and $I + J$ denotes the element being currently processed.

shown. We will first discuss (Section 3.1) the efficiencies obtained with the two methods of reversibly coding the discrete output. Neither the shape of the filter function nor the maximum length which the algorithms can run before a new run is forcibly commenced is varied in the above comparisons. The effect of varying these two parameters is described in Sections 3.2 and 3.3. Some observations are made on free-running algorithms in Section 3.4.

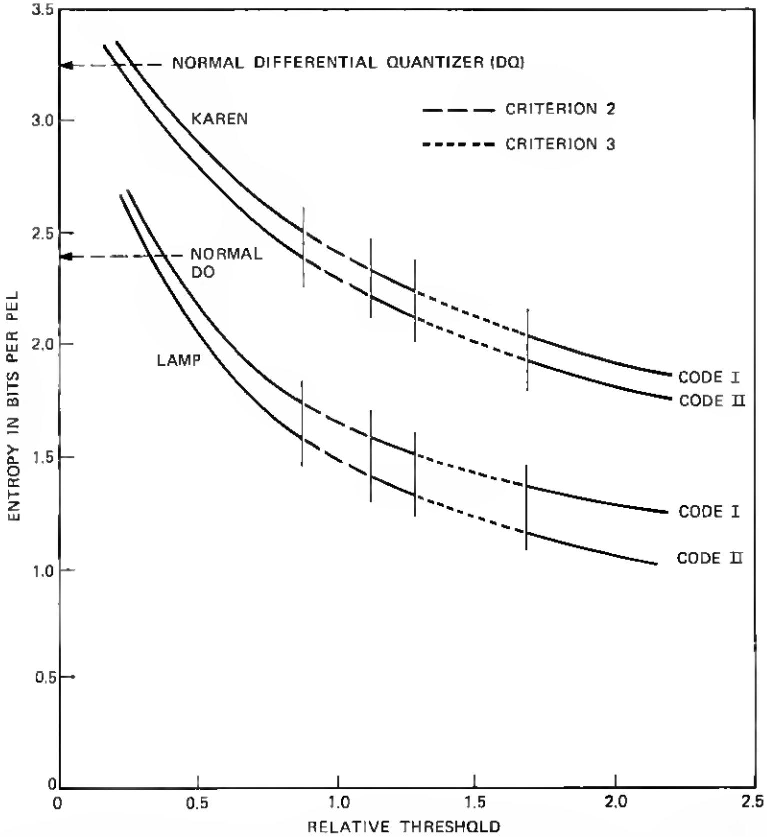

Fig. 7—Entropy of the free-running algorithm as a function of the threshold The measurements of entropy are made under the assumption of two different types of reversible code. The performance for the codes is very similar.

3.1 *Comparison of Reversible Coding Methods*

Figure 7 summarizes the results obtained by applying receiver-model coding interpolatively to the 13-level differential quantizer.* For computational simplicity, the filter used in this case has a rectangular impulse response three elements wide (i.e., corresponding to an average over three elements).

As the threshold is raised on the filtered error sequence, more and more elements are interpolated. Consequently probability distributions become more peaked and the entropy drops. At the same time

* The "relative" threshold is, in fact, one-fifth the threshold value, in 128ths, applied to the filtered error signal.

Fig. 8—(a) Karen—processed by normal 13-level differential quantizer, 1st order entropy 3.10 bits/pel. (b) Picture processed by free-running algorithm, 2.0 bits/pel; picture quality is criterion 3 or worse. (c) Unprocessed picture of "Lamp."

Fig. 8 (continued).

the picture quality is reduced in low-detail areas of the picture as soft detail and texture become blurred. Edges and high-detail areas, however, remain unaffected until very large thresholds are reached.

Consider the results for Karen. As the threshold is increased, the entropy drops from 3.1 bits/pel with a normal differential quantizer* to about 2 bits/pel at which point there is quite noticeable smearing in low-detail areas. Also shown on the curves are the criterion 2 and criterion 3 ranges (Section 2.2). Not until the threshold is raised to a value of 0.9 and the entropy has fallen to 2.4 bits/pel does the change in picture quality become visible when compared with a normal differential quantizer, other than by close A–B comparison. The normal differentially quantized picture is shown in Fig. 8a while the picture coded with a threshold of 1.5 (2.0 bits/pel) is shown in Fig. 8b.

The results obtained with the simpler picture "Lamp" (Fig. 8c) are similar to those obtained for Karen except that the advantage is somewhat greater; the rate is halved in going from the normal dif-

* It is necessary to send additional information to explicitly inform the receiver when to interpolate and when not to. It is this additional information which prevents the entropy of the coded signal from converging to the value of the normal differential quantizer.

ferential quantizer to the end of the criterion 3 range. It is to he expected (see Section V) that low-detail pictures will he more amenahle to receiver-model coding given the present model.

There is surprisingly little difference in efficiency hetween the two reversihle codes, particularly for Karen where the statistics for the highly detailed parts swamp the peaked distributions ohtained in the low-detailed parts. In such instances an adaptive strategy would be of some help.[4] The complexity associated with implementing the simple code (code I) does not change with the maximum permitted length of run; for the variahle code (code II) there is a proportional relationship since a code dictionary would need to he stored for each run position. Consequently, it is important to know how the entropy changes with the maximum length of run that is permitted. For the moment we may conclude that unless the more complex codes can he implemented simply or that channel capacity is at a premium then the simple code is prohahly adequate.

3.2 *Visual Filter Function*

The psychological literature is replete with different estimates of what the shape of the visual point-spread function should he. It was hoped that we could add something to the debate hy investigating different functions in the coding model to see which shape gives the best results. In one experiment the shape of the function was varied keeping the spread of the function constant; the spread was measured hy the first moment of the ahsolute value of the spread function. In a second experiment the spread of the filter was varied keeping the shape constant. Bear in mind that hecause our algorithm works only along the scan-line we cannot take full advantage of the two-dimensional, spatial, point-spread function. Consequently, we should really think of a line-spread function, the rationale being that in the worst-case situation the stimulus heing filtered would have large vertical extent and hence the line-spread function would he appropriate.

3.2.1 *Effect of Shape*

Varying the shape of the filter function has little effect on coding efficiency (Fig. 9). The filter shape was varied from rectangular to quite peaked keeping hoth the area under the function and the first moment of the absolute value of the function constant. The threshold is also constant at 1.0. The square, crosses, and dots of Fig. 9 denote functions with widths of 3, 5, and 7 pels respectively. The values of the functions are given in Table II. For the interpolative algorithm

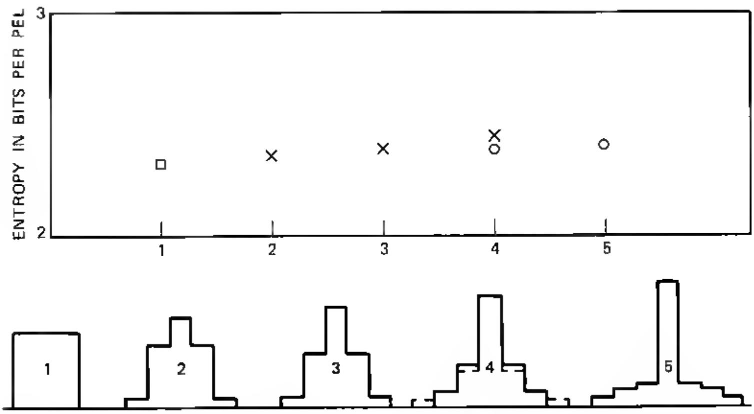

Fig. 9—Effect on entropy (code II) of varying the shape of the filter function. The width of the impulse response is: □—3 elements, ×—5 elements, ○—7 elements. Threshold decisions are very insensitive to the shape of the filter function.

with a maximum runlength of 10 pels there is an increase in bit-rate from 2.32 bits/pel for the rectangular function to 2.41 for the most peaked function; any accompanying change in picture quality was too small to notice.

3.2.2 *Effect of Spread*

The spread of the filter function, on the other hand, has far more effect on the picture quality and entropy than does the shape, as can be seen from Fig. 10a. A rectangular function was used and the spread was varied keeping the area under the impulse response constant and the threshold fixed at 1.0. The picture quality changed from almost

Table II—Weighting Coefficients of Transversal Filter
(The filter shape is symmetrical with A being the central element)

Filter Number	A	B	C	D
1	0.333	0.333	0	0
2	0.4	0.275	0.025	0
3	0.45	0.231	0.044	0
4a	0.5	0.188	0.062	0
4b	0.5	0.156	0.062	0.031
5	0.55	0.103	0.081	0.041

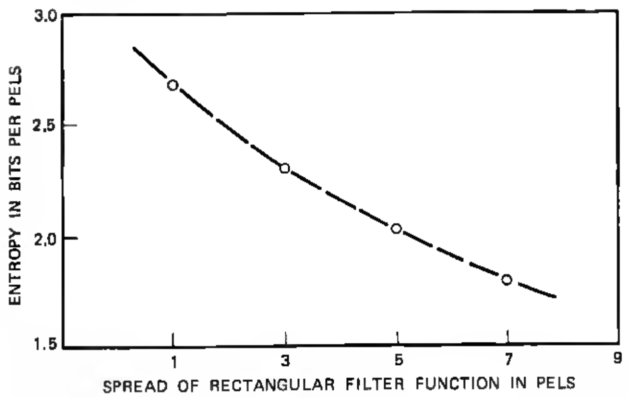

Fig. 10a—The effect on entropy (code II) of varying the width of the filter function. The overall spread of the function has a strong effect on entropy. Subject—Karen.

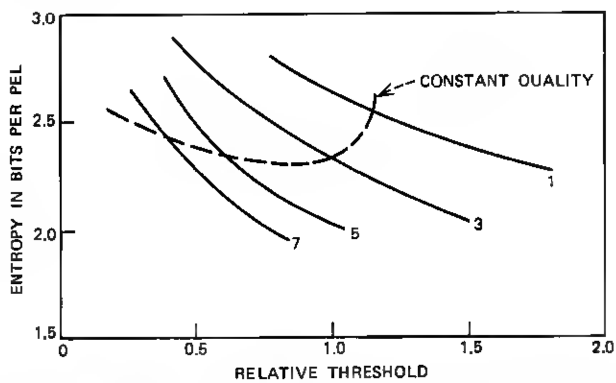

Fig. 10b—Curves of entropy versus threshold for filter functions having spreads of 1, 3, 5, and 7 elements for the free-running interpolative algorithm (Karen). The dashed curve passes through each of the full curves at points of approximately constant picture quality. A spread of between 3 and 5 elements gives the lowest bit-rate for standard viewing distance.

criterion 1 quality with a spread of 1 pel to worse than criterion 3 quality when the spread was 7 pels.

An attempt was made to determine the most suitable filter spread for a picture having criterion 2 quality (standard viewing distance). Figure 10b gives curves of entropy versus threshold for rectangular filter functions of different spread. The dashed curve is a line of approximately constant picture quality. It was determined by making

pair-wise comparisons between a reference picture obtained using a threshold of 1.0 and a filter spread of three and pictures from the other spread curves. With the filter fixed at a particular value of spread the threshold was varied until the picture quality matched that of the reference picture. From the figure it can be seen that a spread of between 3 and 5 pels gives the lowest bit-rate for the standard viewing distance.

3.3 *Effect of Maximum Runlength*

The effect of changing the maximum permitted runlength is shown in Fig. 11. Interestingly, there is very little increase in bit-rate as the maximum runlength is reduced to as little as 4 pels, particularly for code II. Even for the low-detail picture (Lamp) where the average length of a run is much longer, the increase in entropy is still small. Bearing in mind that code II becomes much simpler to implement for short maximum runlengths there appears to be little reason to use long runlengths.

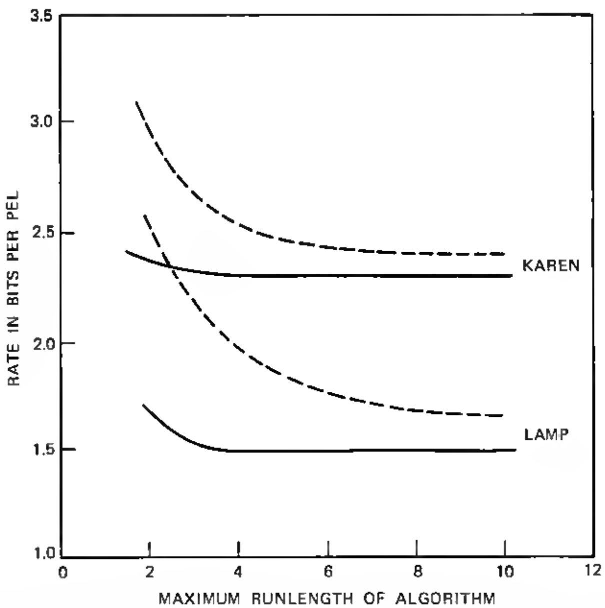

Fig. 11—Entropy as a function of the maximum runlength for Code I (dashed) and Code II (full). Note there is little increase in entropy for Code II as the maximum runlength is reduced from 10 to 4.

Fig. 12—Pictures showing the effect of changing the size of the picture element with the filter function, as measured at the eye, maintained constant: (a) original 8-bit signal, (b) processed, with threshold = 1.5 and H = 1.38 bits/pel, (c) original 8-bit signal, ½ lineal size, (d) processed with same threshold as in (b), H = 1.17 bits/pel. It is the quality difference between pictures of the same size that should be compared, not the relative quality of the two processed pictures.

Fig. 12 (continued).

3.4 *Discussion*

The preceding experiments suggest two ways for decreasing bit-rate at the cost of decreased picture quality. First, it can be decreased by increasing the threshold as shown by Fig. 7. Second, it can be decreased by increasing the spread of the filter function as shown by Fig. 10a. The picture, Karen, was coded to have an entropy (Code II) of 1.81 bits/pel by reducing the quality (lower quality than criterion 3) in the two ways described above. For the first method the filter was rectangular with a spread of three elements while for the second method the filter was again rectangular but with a spread of seven elements. Both methods gave similar picture quality with the narrow-filter/high-threshold combination of the first method being, perhaps, slightly better. The improvement in sharpness of the first method was partly offset by the reduction in granularity and blotchyness of the second method.

If a particular filter, at normal viewing distance, produces a picture that is just distinguishable from a high-quality original then doubling the spread of the filter function should produce a picture at twice the viewing distance which is again just distinguishable from the original.

I have tried to demonstrate this prediction with Fig. 12 by reproducing a comparison pair of pictures at half-size to correspond to the situation where the viewing distance is doubled. It is the *difference* in quality between pairs of pictures at the same viewing distance that should be compared, not the comparative quality of the processed pictures.

One factor that could upset such a comparison is that the smaller picture has a greater scanning line density. The filter function operates in one dimension only and to the extent that deleted picture components are uncorrelated from line to line, vertical filtering taking place in the eye will tend to favor the smaller picture. An intuitive feel

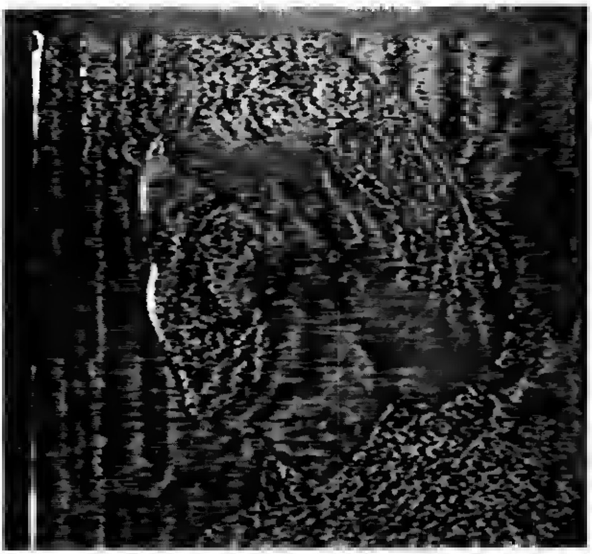

Fig. 13—Picture of the filtered error signal for the processed picture of Fig. 12b.

for the correlated nature of the error signal is obtained from Fig. 13, in which a certain amount of picture structure is evident.

IV. RECEIVER-MODEL CODING WITH GRID ALGORITHMS

4.1 *Introduction*

One can take advantage of the filtering action of vision without explicitly filtering the error signal. To appreciate this, let us consider the following grid algorithm. Every grid element (sampled point) is reproduced with full accuracy (e.g., 7 or 8 bits). The intermediate elements (referred to as "conditional points") are reproduced as the average of the adjacent pels, $\bar{X}_{i+1} = (X_i + X_{i+2})/2$, if the error $(\bar{X}_{i+1} - X_{i+1})$ is small (see Fig. 14). Otherwise, the error quantity is quantized and transmitted. In determining whether $\bar{X}_{i+1}$ is an adequate representation of X_{i+1}, the error signal adjacent to pel $(i + 1)$ must be filtered. However, the error at pels i and $(i + 2)$ is virtually zero so that for a filter that consists of a three-point average it is only necessary to examine the error introduced at pel $(i + 1)$.

Kretzmer[24] proposed a coding scheme similar to the above in which every fourth pel is always coded with 7-bit accuracy (i.e., 4:1 grid algorithm). The intermediate points are estimated by linear inter-

polation and the difference between the input and the estimate is quantized and transmitted. The midpoint in each quad is quantized more accurately than the quarter and three-quarter points. Fukushima and Ando[25] experimented with a very similar scheme in which every fourth point was transmitted with 6-bit accuracy and the intermediate points were transmitted using three levels. A final bit-rate of 2.7 bits/pel was achieved. They also investigated two-dimensional 4:1 algorithms. Connor has investigated a 2:1 grid algorithm (column coder) which uses two-dimensional prediction for differentially coding the grid points.[26] Pease[27] has applied what amounts to a 2:1 grid algorithm between fields of a television picture. All points in one field are estimated as the average of the four surrounding points coming from the previous and next fields. Only when this prediction breaks down is additional information sent about the interpolated field. In the presence of movement the four-way interpolation is less accurate and the numher of pels that require correction increases somewhat. Notice that all the ahove schemes transmit two or more different types of amplitude information; the grid points are transmitted ahsolutely (or differentially, relative to one another) while the conditional points are transmitted as a *correction* to the estimation. These schemes will therefore be referred to as error transmission schemes.

In this section we will examine a numher of grid coding schemes. For the most part they differ from the above schemes in that only one type of amplitude signal is transmitted so that all amplitude information is decoded in the same way (direct transmission). The distinction is best appreciated hy considering a specific example. Take the inter-

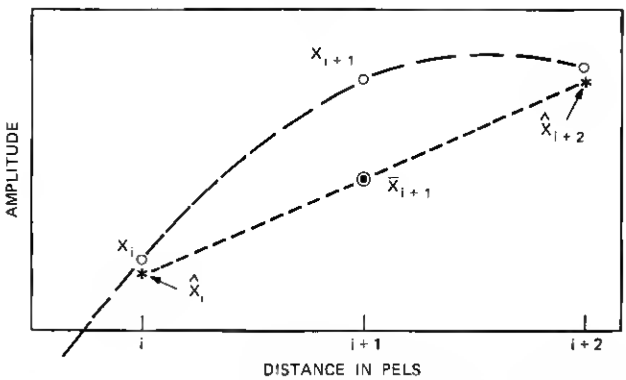

Fig. 14—Definition of locations and values of elements used in discussion of grid algorithms.

polative algorithm: if pel i has already heen encoded (Fig. 14), pel $(i + 2)$ is then encoded differentially from pel i. From the encoded values of pel i and pel $(i + 2)$, $(\hat{X}_i, \hat{X}_{i+2})$, $\bar{X}_{i+1}$ is formed. The error signal $(X_{i+1} - \bar{X}_{i+1})$ is tested against the threshold; if it exceeds threshold, pel $(i + 1)$ is differentially coded from pel i and pel $(i + 2)$ is differentially recoded from pel $(i + 1)$. Thus it can he seen that the interpolated value $\bar{X}_{i+1}$ is only retained when the interpolation is adequate; otherwise it is discarded. Furthermore, the quantizing scales for pels i and $(i + 1)$ can he the same as for normal differential quantization since in high-detail areas the interpolation generally fails and each element is predicted from the previous element. In practice, a check is made to determine whether slope overload will occur in coding pel $(i + 2)$; if this can happen pel $(i + 1)$ is then coded and pel $(i + 2)$ is recoded, differentially, from pel $(i + 1)$. Thus, in high-detail parts of the picture, pel $(i + 1)$ is rarely interpolated and the coding operation differs little from normal differential quantization. In low-detail parts of the picture, where the interpolation process is usually adequate, again the coding process is normal differential quantization, hut with twice the normal sample spacing.[4]

Errors will occur at pels i and $(i + 2)$ hecause differential quantization has been used and these errors will, because of the visual filtering action, affect the visihility or the error occurring at pel $(i + 1)$. Hence the encoding will he more efficient if filtering is used. But, as we will see, a three-point filter does not differ much from a single-point filter hecause the errors made at pels i and $(i + 2)$ are limited hy the numher and spacing of the quantizer levels and cannot he suhjectively large if adequate quality is to he ohtained.

In comparing the error transmission and direct transmission schemes, it can he seen that the decision on whether or not to transmit the conditional elements is the same in hoth cases. The error transmission scheme has the advantage that the estimate is a hetter prediction than the previous sample, and hence the correction signal, where it is necessary to transmit it, will he smaller. However, the disadvantage is that since the grid points are transmitted as differences from a point two pels away, the amplitude of the differences and hence the entropy associated with them will he larger. In practice this will increase complexity since the quantizer will need to have more levels to handle the larger changes. In Section 4.2 an error transmission scheme will be compared with a number of direct transmission algorithms and it will he seen that there is very little difference in performance hetween the two types of schemes. One would expect the

performance to converge for low-detail pictures since the number of points which are not successfully interpolated becomes very small and the encoding of the remaining points is then very similar.

In the free-running algorithms a special code word was used to inform the receiver when to interpolate. For the grid algorithms an interpolate command has been inserted in a special manner. On the conditional samples only, the zero differential quantizer level is used to denote the interpolate command: this means that when the signal is not being interpolated the zero level cannot be used; instead the signal is forced to take on the next closest level, either the positive or negative inner level. This affects picture quality very little since, firstly, a zero level is rarely used on the conditional samples and, secondly, since interpolation generally fails in the vicinity of large luminance changes, the small error introduced by deleting the zero level is largely masked by the consequent luminance change.

Implementation of the grid algorithm becomes even simpler when we consider two variations, a modified form of the interpolative (MI) algorithm and an extrapolative algorithm. The MI algorithm is quite similar to the interpolative algorithm; the next grid point is *not* quantized prior to interpolation. This means that it is only necessary to quantize each element sequentially just as one does in normal differential quantization (when a pel is adequately interpolated, the classifier output is simply forced to a zero prior to processing by the local [and distant] decoder and the next element [pel $i + 2$] is processed in the normal manner [see Fig. 14]). In the extrapolative algorithm the method used to estimate the conditional sample is the same as the method of extrapolation for the coding process (i.e., previous sample prediction) and hence the need for an extrapolate command is obviated. The algorithm is then only slightly different from normal quantization, especially if the error occurring at the conditional sample is taken as the filtered value (the scheme described in Ref. 4 under the name "Level Variable Sampling Scheme").

4.2 *Comparison of Free-Running and Grid Algorithms*

The performance of both a 2:1 and a 4:1 MI, grid algorithm are compared with the free-running extrapolative algorithm in Fig. 15.

The maximum reduction that can be obtained with the 2:1 algorithm is a halving of the bit rate. Long before this point is reached the curve starts to flatten out and unless very large thresholds are used the picture quality remains high. Within the obtainable range of picture quality the 2:1 algorithm performs almost as well as the free-running

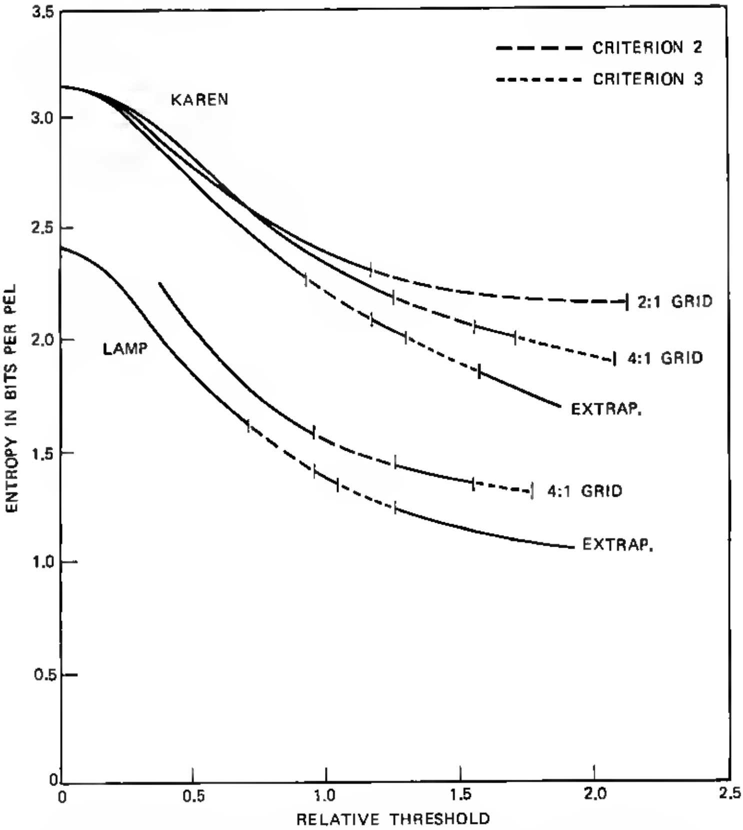

Fig. 15—Comparison of the performance of free-running and grid algorithms. The 4:1 grid algorithm performs equally as well as the free-running algorithm.

algorithm. By going to the 4:1 algorithm, a larger picture quality range can be accommodated without going to very large thresholds. In the criterion 2 range the grid algorithm seems slightly better than the extrapolative free-running algorithm while in the criterion 3 range the free-running algorithm is marginally better.

4.3 *Comparison of Three Grid Algorithms—Error-transmission, MI, and Extrapolative*

Since the MI and error-transmission algorithms are the most alike, we will compare them first. The error-transmission algorithm uses a 19-level differential quantizer. This is obtained from the 13-level

quantizer by adding additional outer levels. The filtered error signal is obtained by summing the error at the estimation point and the two adjacent grid points. The MI algorithm uses the usual 13-level quantizer and the filtered error signal is the sum of only two error terms. The quantizing error occurring at the grid point to the right of the point being interpolated cannot be included since this point is not quantized until after a decision has been made on the conditional point.

The white markers in Fig. 16 indicate those conditional points in the two algorithms for which the filtered error signal is above threshold. Hence these points are not adequently represented by the estimate (the relative threshold is set at 1.5 for both algorithms). The distribution of markers is quite similar, especially when one bears in mind that the error summing procedure is different in the two cases. The picture quality and hit-rate is also very similar (see Fig. 17), which stands to reason since the signal is processed identically in those parts of the picture where there are no markers. The algorithms were evaluated on other pictures. In each case picture quality and hit-rate were very close.

The extrapolative (like the MI) algorithm uses a 13-level quantizer and sums the error over only two pels. The estimation procedure (zero-order-hold) is not as effective as linear interpolation and, as a result, the number of conditional points that need to be transmitted is very much larger for a specific threshold. A consequence is that the curve of entropy versus threshold lies above the other curves except at higher thresholds. Here, the curves converge since the only conditional points still being transmitted are edge points. The picture quality is not quite as high as that obtained with the other two algorithms with the defect appearing as a granularity in flat, dark regions of the picture. Although the granularity is also present for the other two algorithms it is significantly attenuated by the interpolative averaging.

4.4 *Effect of Filtering*

As indicated previously, the effect of filtering for the 2:1 grid algorithm will not be very strong since when the error is evaluated at each conditional point the error permitted at the adjacent points, which are quantized with full accuracy, will be quite small. Even so, there is a small increase in the number of conditional flags that are transmitted in going from the single-point filtering to the two-point

Fig. 16—Markers showing conditional points that were updated with a threshold of 1.5: (a) error transmission algorithm, (b) MI algorithm.

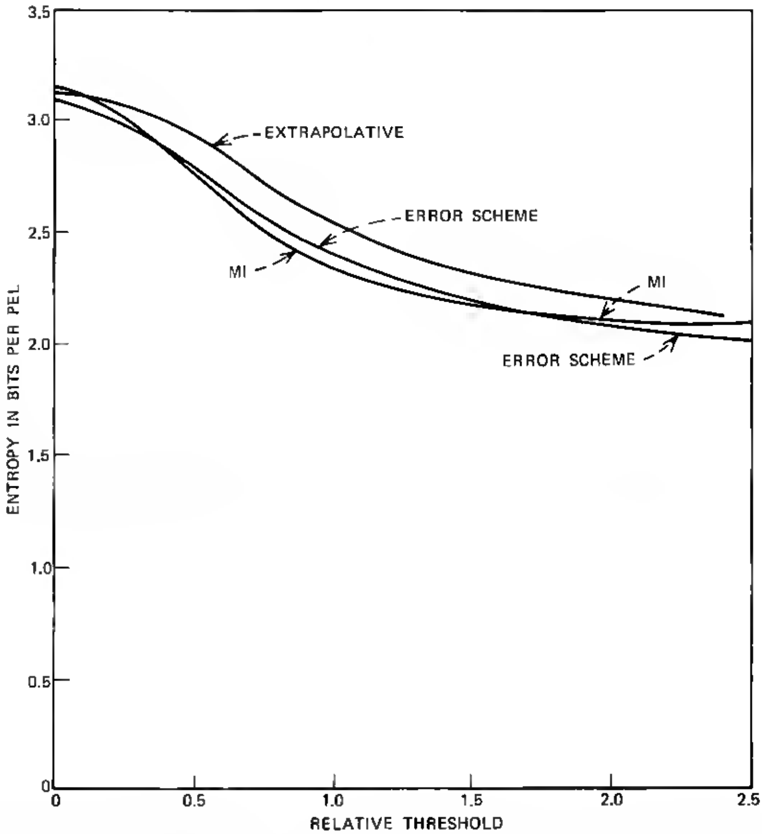

Fig. 17—Relative performance of three different 2:1 grid algorithms. The extrapolative algorithm is slightly inferior to the MI and dual-mode algorithms. Subject—Karen.

filtering (error at conditional point plus the error at previous point). This, in turn, results in a small increase in entropy (from 2.17 to 2.20 bits/pel).

For the 4:1 fixed-point algorithm the difference between single-point and three-point filtering is larger. The conditional points that are transmitted have heen marked in Fig. 18 where, for single-point filtering, the threshold is 0.9 and the entropy is 2.08 bits/pel and for three-point filtering the threshold is 1.5 and the entropy is 2.04 bits/pel. In this case, however, the effect on picture quality is more noticeable. With the broader filter low-detail areas are reproduced hetter while medium-detail areas appear more noisy. At normal viewing dis-

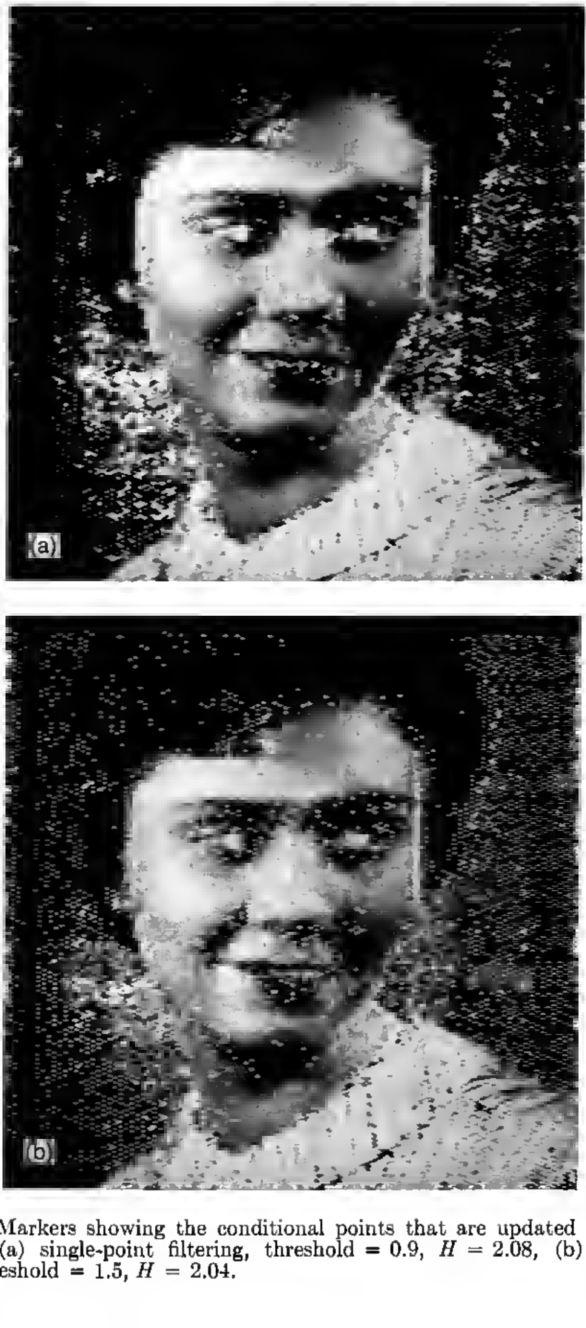

Fig. 18—Markers showing the conditional points that are updated for 4:1 grid algorithm: (a) single-point filtering, threshold = 0.9, H = 2.08, (b) three-point filtering, threshold = 1.5, H = 2.04.

tances the broad filter is preferable while for close scrutiny the single-point filter is better.

There is no reason why filtering could not take place in two or three dimensions in which case more elements would be involved and the accuracy with which the picture was encoded could more accurately match perceptual requirements for a given viewing situation.

V. DISCUSSION

As we have seen, the receiver-model coding algorithm with the simple threshold model of Fig. 2 tends to work best on low-detailed pictures. There are two reasons for this: (*i*) In detailed parts of the picture the estimation procedure is not as good as in low-detail areas; (*ii*) The threshold model, as described, is a simple, low-pass filter model and does not incorporate the effects of masking by adjacent signal components such as occurs when an element lies close to a large change (spatially or temporally) in luminance.*

The receiver-model coding concept, as stated, does not depend on any specific receiver model. As better models of the human viewer are obtained they can be incorporated directly into the encoding operation. In essence it is a three-step operation: estimation, testing, and, if necessary, more accurate recoding. There is an intrinsic separation between the source-property operation (estimation) and the receiver-property operation (testing) and as such the technique will be suboptimum. Performance could undoubtedly be improved by cycling through the estimate-test-recode sequence iteratively.[28] The interesting, practical question would be, is the improved performance worth the added complexity?

In all the coders described here the bit-rate—picture-quality operating point is determined by means of a single threshold control. This means that it is a relatively simple matter to dynamically alter the operating point in response to some system requirement. An example occurs in frame-to-frame coding where the moving area is transmitted as an element-differentially-quantized signal. As the buffer fills in response to increased movement the threshold is raised so as to keep the data-generation rate more uniform.[29]

VI. SUMMARY AND CONCLUSIONS

Receiver-model coding is a powerful, though not optimal, technique for incorporating properties of the human observer into the picture

* Some practical coding strategies have been developed that take advantage of spatial masking effects.[4,6]

encoding process. In essence, components of the signal are estimated according to some algorithm. The difference between the actual signal and the estimate is processed in a model of the receiver to determine if the estimate is adequate. If so, the receiver is informed of this; if not, additional information is transmitted to improve the estimate.

The receiver-model coding concept may be applied in many different ways and the visual model may range from very simple to very complex. In this paper I have used the differential quantizer (DPCM coder) as the basic vehicle with which to investigate receiver-model coding, and the visual model is a one-dimensional low-pass filter. Three types of estimation are investigated: extrapolation, interpolation, and a simplified form of interpolation referred to as "modified interpolation." It is important to bear in mind that the estimation is used to help determine *which* components need to be transmitted and does not indicate *how* the components are transmitted. In nearly all examples considered here the transmitted component is a simple difference signal which is decoded by adding the difference to the last decoded value.

Coders are divided into two separate classes, free-running algorithms and grid algorithms. In the free-running algorithms the estimation procedure may continue in a single run until the estimate fails with the proviso that the length of the run may not exceed a specified maximum. With the grid algorithm a fixed set of elements is always transmitted (e.g., every second or every fourth element). The interest in grid algorithms stems from the fact that they are more easily implemented.

The free-running interpolative algorithm gives a reduction in entropy of approximately 30 percent for high-detail pictures and 50 percent for low-detail pictures for a small loss in picture quality when the picture is evaluated by observing a single "frozen" frame on a high-quality CRT display.

Two reversible coding strategies were explored for converting the quantizer output to a binary code. Code II gives an advantage of between 0.1 and 0.15 bits/pel over Code I when using a maximum runlength of ten elements; the relative advantage of Code II over Code I about doubles when the maximum runlength is reduced to four elements.

The effect of the threshold filter function on the coding operation was explored by varying the shape of the filter function while keeping the spread of the function constant and then, in a second experiment, keeping the shape constant and varying the amount of spread. While

the exact shape of the filter function affected performance very little, the spread of the function had a large effect; the most suitable spread appears to be about three elements for the normal viewing distance.

As the maximum permitted length of run is decreased from 10, it is found that there is very little increase in entropy for Code II for a maximum runlength even as short as 4, suggesting that a 4:1 grid algorithm may perform almost as well as free-running algorithms.

The 2:1 grid algorithm (modified interpolative) does not permit operation at lower picture qualities and bit rates; the 4:1 algorithm has a larger range. However, over their range of operation, the grid algorithms perform at least as well as the best free-running algorithm and in view of their simpler implementation appear to be the most promising.

Three different 2:1 grid algorithms were compared, an error-transmission technique in which the correction signal is sent as a difference between the estimate and the input, the modified interpolative algorithm, and the extrapolative algorithm. Extrapolation was slightly inferior to the other two methods and of these the modified interpolative method is more simply implemented.

The emphasis in this paper has been on obtaining an efficient discrete representation of a picture signal rather than presenting a complete coding system. Consequently, there are a number of considerations such as sensitivity to transmission errors which are not discussed in the paper but nevertheless bear importantly on the feasibility of any practical coder.

VII. ACKNOWLEDGMENTS

This study would have been extraordinarily difficult without the picture processing facility, both hardware and user-oriented system software, of R. C. Brainard and J. D. Beyer. I thank them for their gentle guidance.

REFERENCES

1. Graham, R. E., "Subjective Experiments in Visual Communication," IRE Conv. Rec., 1958, pp. 100–106.
2. Powers, K. H., and Staras, H., "Some Relations Between Television Picture Redundancy and Bandwidth Requirements," Trans. Amer. IEE, *76* (1957), p. 492.
3. Graham, R. E., "Predictive Quantizing of Television Signals," IRE Wescon Conv. Rec., Part 4, 1958, pp. 142–157.
4. Limb, J. O., "Adaptive Encoding of Picture Signals," in Huang, T. S., and Tretiak, O. J., eds., *Picture Bandwidth Compression*, Gordon and Breach, Science Publishers, 1972.
5. Wintz, P. A., and Tasto, M., "Picture Bandwidth Compression by Adaptive Block Quantization," Purdue University School of Electrical Engineering, TR-EE 70-14, July 1970.

6. Brown, E. F., and Kaminski, W., "An Edge-Adaptive Three-Bit Ten-Level Differential PCM Coder for Television," IEEE Trans. Commun. Tech., *COM-19*, No. 6 (December 1971), pp. 944–947.
7. Limb, J. O., "Source-Receiver Encoding of Television Signals," Proc. IEEE, *55*, March 1967, pp. 364–379.
8. Candy, J. C., and Bosworth, R. H., "Metbods for Designing Differential Quantizers Based on Subjective Evaluations of Edge Busyness," B.S.T.J., *51*, No. 7 (September 1972), pp. 1495–1516.
9. Budrikis, Z. L., "Visual Fidelity Criterion and Modeling," Proc. IEEE, *60*, No. 7 (July 1972), pp. 771–779.
10. Kortman, C. M., "Redundancy Reduction—A Practical Method of Data Compression," Proc. IEEE, *55*, No. 3 (March 1967), pp. 253–263.
11. Novak, S., and Sperling, G., "Visual Thresholds Near a Continuously Visible or Briefly Presented Light-Dark Boundary," Optical Acta, *10*, No. 2 (April 1963), pp. 87–91.
12. Fiorentini, A., "Mach Band Phenomena," in Jameson, D., and Hurvich, L. M., eds., *Handbook of Sensory Physiology*, Vol. VII/4, Springer-Verlag, 1972.
13. Limb, J. O., "Vision Oriented Coding of Visual Signals," Ph.D. Thesis, University of Western Australia, 1966.
14. Moon, P., and Spencer, D. E., "The Visual Effect of Non-Uniform Surrounds," J. Opt. Soc. Amer., *35*, March 1945, pp. 233–248.
15. Hacking, K., "Tbe Relative Visibility of Random Noise Over the Grey-Scale," J. Brit. IRE, *23*, No. 4 (April 1962), p. 307.
16. Newell, G. F., and Geddes, W. K. E., "Visibility of Small Luminance Perturbations in Television Displays," BBC Research Department, Report T106, 1963.
17. Budrikis, Z. L., "Visual Threshold and the Visibility of Random Noise in TV," Proc. IRE Australia, *22*, December 1961, pp. 751–759.
18. Blackwell, H. R., "Neural Theories of Simple Visual Discriminations," J. Opt. Soc. Amer., *53*, January 1963, pp. 129–160.
19. Kristofferson, A. B., "Visual Detection as Influenced by Target Forms," in Wulfeck, J. W., and Taylor, J. H., eds., *Form Discrimination as Related to Military Problems*, National Academy of Sciences—National Researcb Council Publication, 561, 1957, pp. 109–126.
20. Budrikis, Z. L., "Model Approximations to Visual Spatio-Temporal Sine-Wave Threshold Data," to be published in November 1973 B.S.T.J.
21. Mounts, F. W., and Pearson, D. E., "Measurements of the Apparent Increase in Noise Level Resulting from Frame-Repetition of Low-Resolution TV Pictures," B.S.T.J., *48*, No. 3 (March 1969), pp. 527–539.
22. Limb, J. O., "Efficiency of Variable-length Binary Codes," Proc. Univ. Missouri, Rolla, M. J. Kelly Communications Conf., 1970, pp. 13-3-1, 13-3-9.
23. Rice, R. F., and Plaunt, J. R., "Adaptive Variable-length Coding for Efficient Compression of Spacecraft Television Data," IEEE Trans. Commun. Tech., *COM-19*, No. 6 (December 1971), pp. 889–897.
24. Kretzmer, E. R., "Reduced Bandwidth Transmission System," U. S. Patent No. 2,949,505, August 16, 1960.
25. Fukushima, K., and Ando, H., "Television Band Compression by Multimode Interpolation," Technical Research Laboratories, Japan Broadcasting Corporation, Japan.
26. Connor, D. J., "Techniques for Reducing the Visibility of Tranmission Errors in Digitally Encoded Video Signals," IEEE Trans. Commun. Tech., *COM-21*, No. 3 (June 1973).
27. Pease, R. F. W., "Conditional Vertical Subsampling—A Technique to Assist in the Coding of Television Signals," B.S.T.J., *51*, No. 4 (April 1972), pp. 787–802.
28. Viterbi, A. J., "Convolutional Codes and Their Performance in Communication Systems," IEEE Trans. Commun. Tecb., *COM-19*, No. 6 (October 1971), pp. 751–772.
29. Limb, J. O., Pease, R. F. W., and Walsh, K. A., "Combining Intraframe and Frame-to-frame Coding for Television," unpublished work.